DOCUMENT RESUME

ED 140 261 CS 003 492

AUTHOR Fischer, Joan, Ed.
TITLE 9 Questions on New Jersey State Testing on Reading; Proceedings of the Spring 1975 Rutgers University Reading Conference.
INSTITUTION Rutgers, The State Univ., New Brunswick, N.J. Reading Center.; Rutgers, The State Univ., New Brunswick, N.J. Univ. Extension Div.
PUB DATE 75
NOTE 71p.; Paper by Dyer p. 1-5, and by Lennon, p. 50-52 removed due to copyright restrictions

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS Conference Reports; *Educational Testing; Elementary Secondary Education; *Reading; Reading Diagnosis; *Reading Tests; *State School District Relationship; *State Standards; *State Surveys; Student Evaluation
IDENTIFIERS *New Jersey

ABSTRACT

At a reading conference devoted to analysis of the impact of statewide testing of reading in New Jersey, the following nine questions were addressed: How can scores be used by teachers? How should test results be reported to the press and the public? Do state tests replace commercial, normed tests? Is the state test unfair to minorities? How might test scores change the present curriculum? Are the scores diagnostic? Can teaching of reading be improved by using the state test? Why are scores given so that comparisons may be made between schools in a district and between districts? What changes could or should be made in the tests to make them more useful? These questions stimulated discussions of priorities in statewide evaluation, of using assessment information in educational planning and decision making, of guidelines for interpretation of assessment program and results, of the formation of local norms, and so on. Test specifications and reading-test results for grades four, seven and ten are described. (KS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED140261

RUTGERS UNIVERSITY
the state university of new jerse
university extension division
conferences department
in cooperation with
reading center, graduate
school of education

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

9 Questions ON New Jersey State Testing ON Reading

CS803 472

Edited by
JOAN FISCHER
Rutgers Reading Center and
The Adult Education Project

THE 9 QUESTIONS

General Session Speakers, Section Chairmen and Panel Members will answer the following questions with any emphasis that they choose. They may give very brief answers to some and elaborate on others at their option. At the Section Meeting, members of the audience may ask any further questions.

1. How can the scores be used by teachers?
2. How do you report test results to the press and the public?
3. Do state tests replace commercial normed tests?
4. Is the state test unfair to minorities?
5. How might test scores change the present curriculum?
6. Are the scores diagnostic?
7. Can teaching of reading be improved using the state test?
8. Why are scores given so that you can make comparisons between schools in a district and between different districts?
9. What changes could or should be made in the tests to make them more useful?

PROGRAM

8:15 a.m. REGISTRATION OPENS--COFFEE AVAILABLE

CHAIRMAN:
Dr. Edward Fry
Director, Reading Center
Rutgers University

INTRODUCTION:
Dr. Edward Bloustein
President, Rutgers University

9:00 a.m. GENERAL SESSION

Dr. Roger Lennon
Senior Vice President
Harcourt Brace Jovanovich
Former N.J. State Test Contractor

Dr. Roy Kress
Professor of Psychology of Reading and Acting Dean
College of Education
Temple University

Dr. Henry S. Dyer
Former Vice President
Educational Testing Service

Dr. Robert Thorndike
Professor, Teachers College of Columbia University
President, American Educational Research Association

TABLE OF CONTENTS

FOREWORD, Edward Fry. i

STATEWIDE EVALUATION--WHAT ARE THE PRIORITIES
Henry S. Dyer. 1

UTILIZING ASSESSMENT INFORMATION IN EDUCATIONAL PLANNING AND DECISION-MAKING
Gordon Ascher. 6

TEST SPECIFICATIONS, Grade 4. 25

TEST SPECIFICATIONS, Grade 7. 28

TEST SPECIFICATIONS, Grade 10 30

READING TEST RESULTS: Grade 4 32

Grade 7 34

Grade 10. 37

THE BIG PICTURE: HOW MANY PUPILS AND TEACHERS ARE THERE IN NEW JERSEY. 41

GUIDELINES FOR THE INTERPRETATION OF THE NEW JERSEY EDUCATIONAL ASSESSMENT PROGRAM RESULTS 42

TO PERFORM AND TO ACCOUNT
Roger T. Lennon. 50

IT'S EASY TO MAKE LOCAL NORMS
Edward Fry 53

USING GROUP READING ACHIEVEMENT TESTS TO IDENTIFY PROBLEM READERS
Roy A. Kress 57

A WORD TO TEACHERS ON STATEWIDE TESTING IN NEW JERSEY
Margaret Smith-Burke 61

READING AS REASONING
Robert L. Thorndike. 65

Foreword

by

Edward Fry
Conference Chairman

Putting together a conference on reading is a lot of work as anyone who has tried it will testify. Putting together this one was more than usual because we tried valiantly to include the State Department of Education in the planning.

Unfortunately, just before we were ready to go to press with the announcements, the State Department informed us that they decided not to participate and withdrew their staff as speakers. According to a letter from Assistant Commissioner Shine to Dr. Victor Podesta, his designated coordinator, "This action was made necessary by recent directives from the Budget Bureau to reduce our current level of expenditures for the remainder of this fiscal year. Services not directly associated with decentralization and thorough and efficient are being reduced." It is a little hard for me to see how the budget problems can be solved by not sending speakers to this conference, but then I wouldn't be the first one not to understand the bureaucratic mind.

It might be worth noting that the State pays about $1.27 per pupil per year to conduct the statewide tests in reading and mathematics. This amount covers the testing and reporting the results on every 4th, 7th, and 10th grader in New Jersey public schools. There are approximately 100,000 students in each grade. This figure does not include any time, effort, and money expended by local school districts which, in some cases, is considerable.

With such a substantial financial and human effort put into this testing program, we feel that it is worthwhile to ask our 9 questions and such others as might arise in the minds of the audience.

In 1972 the State Board of Education put out policy guidelines for the Educational Assessment Program of which the current testing is a part. This policy states that "since the general purpose of an assessment program is the development interpretation and dissemination of information, all information developed will be made available to all interested parties within the framework of laws of New Jersey and the State Board policies." This policy also holds that "assessment results that can be related to individual teachers shall not be publicly divulged" and "Assessment results that can be related to an individual pupil shall be divulged only to that pupil, his parent or legal guardian, and teachers."

We have a massive amount of group data here which should be studied and wherever possible, utilized. We hope that this conference will stimulate interest in this data and the ongoing testing program. There are many roundhouse criticisms of the teaching of reading in this and most other states. Perhaps the current testing program or a future testing program with modifications can give us some data to either refute or correct criticisms. Perhaps the data will also help the schools to focus their efforts on the two fundamental subjects of reading and mathematics, rather than have them diverted by transient or special interests.

I have not seen any indication which could be used to to evaluate the reading level of N.J. students or data that can be used over succeeding years to see if any overall progress is being made. With only a little modification in the testing program and in the statistical analyses, this type of information could be made available to the public.

Understanding testing can be complicated. While efforts have been made to simplify the reporting of the state tests, still greater knowledge on the part of educators will give them better utilization of the present results and perhaps enable them to ask for improvements. Test results that are given item by item as these are still are only one way to report results; total scores or subtest scores might be considered as a next step.

Those two old standards of testing reliability and validity need to be continually examined. Content validity needs to be continually reexamined to see if the tests are really measuring the kind of skills that are important. Reliability is a major problem in item by item scoring -- how reliable is a one item test? Would a student score the same if he took the item again or took a similar item on the same content? When you don't use total scores, you have a difficult, if not impossible, time of establishing reliability; yet teachers are expected to develop curriculum needs from single items -- perhaps.

We hope that this conference will enable educators to use not only the state tests but all tests a bit more effectively. It is possible that it might stimulate them to ask better questions of their local experts or send them to review their notes from a Tests and Measurements course.

We should keep in mind that tests are made not to harm or penalize students, but to help them by giving objective information and assisting the teacher and administrators in better educational planning.

STATEWIDE EVALUATION--WHAT ARE THE PRIORITIES? by Henry S. Dyer Phi Delta Kappan, June, 1970.

Pages 1-5 were removed due to copyright restrictions.

UTILIZING ASSESSMENT INFORMATION IN EDUCATIONAL PLANNING AND DECISION-MAKING*

Gordon Ascher
State Department of Education
Trenton, New Jersey

The New Jersey Educational Assessment Program is designed to provide useful information to educational decision-makers who operate on a variety of levels. Demands for information come from a variety of State level and local level sources. Governor William T. Cahill in his State of the State address delivered on January 11, 1972 said:

> For although we have made significent strides in recent years toward improving the efficiency and quality of education, the State continues to experience difficulty in accurately determining the area of greatest need. We know some disparities in achievement result in inequities in the distribution of wealth from district to district but we cannot pinpoint the relationsnip between expenditure and achievement. In view of recent movements toward equalizing educational opportunities throughout the State, particularly in the courts, it is essential that current and accurate data be made available.
>
> I am alarmed, quite frankly, about reports of lack of reading ability on the part of children, particularly in the urban areas; and I am concerned that there is no reliable scientific test on a Statewide basis to determine reading ability and reading growth of our youth (Cahill, 1972).

The State Assembly passed a bill which mandates a testing program. The supporters of the bill expressed a need to find out where educational efforts should be concentrated. The bill which was not

*Major portions of this paper were incorporated into a paper read at the American Educational Research Association convention, New Orleans, Louisiana, February 27, 1973 entitled, "Individualized Instruction and State-wide assessment."

acted upon by the Senate, would have amended an existing statute in that it required "uniform Statewide tests and examinations of achievement in reading and mathematics . . . ." (New Jersey State Assembly, 1972).

The New Jersey Constitution gives the Legislature the responsibility of providing a proper education for the State's children. New Jersey statutes provide to the Commissioner of Education the authority, and responsibility, to determine the "thoroughness and efficiency" of education. New Jersey Statute 18A:4-24 states that:

> The Commissioner shall, by direction or with the approval of the state board, whenever it is deemed to be advisable so to do, inquire into and ascertain the thoroughness and efficiency of operation of any of the schools of the public school system of the state and of any grades therein by such means, tests and examinations as to him seem proper, and he shall report to the state board the results of such inquiries and such other information with regard thereto as the state board may require or as he shall deem proper, but nothing in this section shall affect the right of each district to prescribe its own rules for promotion.

Dr. Carl L. Marburger, past Commissioner of Education, provided the following introduction to a Question and Answer brochure designed to inform professional educators and the public about the Educational Assessment Program:

> This fall, New Jersey's educational system will embark on a new venture, the Educational Assessment Program. The purpose of this program is to assess the status of education in New Jersey relative to the statewide educational goals adopted by the State Board of Education earlier this year.
>
> We in education are being called upon to verify that the schooling provided for our students is effective and meaningful, and to demonstrate that the investment of time and financial resources results in quality education. Increasingly, the Governor and the Legislature are seeking such data as a basis for decisions concerning education. Clearly, New Jersey is faced with a mandate, exemplified by recent decisions and recommendations of the courts and the Governor's Tax Policy Committee, to develop an appropriate assessment program for our educational system in New Jersey. We not only need to know more precisely what our educational

problems are across the State, but also we need insight into how to improve educational conditions and how to identify outstanding programs that could be applied in other school districts. As a means of answering concerns such as these, the Educational Assessment Program has been developed by the State Department of Education (New Jersey State Department of Education, 1972 a).

The court decision referred to by Marburger is a decision delivered by Judge Theodore I. Botter on January 19, 1972 in which he indicated that the Commissioner should exercise his power to use tests to determine educational needs. This decision has been upheld by the New Jersey Supreme Court.

At the same time that State level personnel were calling for additional information, local level personnel were moving toward more individualized instruction and were developing needs for individualized assessment. This movement was supported by many programs emanating from the State Department of Education. What Stufflebeam, et al, (1971) called "the problem of administrative levels" was becoming acute in New Jersey. The Educational Assessment Program was designed to provide information on a number of levels.

GOALS

The Educational Assessment Program goals which relate to State-level information needs are:

> To assess the status of the New Jersey educational system relative to the Statewide goals as determined by the State Board of Education through the "Our Schools" Project.
>
> To provide information useful in the planning and evaluation of educational programs designed to bring the New Jersey educational system closer to those Statewide goals.

The goals which relate to the needs of both State-level and local-level personnel are:

> To identify local school districts which, among others sharing similar circumstances, have conducted programs which have been most successful in achieving stated goals.
>
> To identify local school districts in need of State

Department of Education assistance.

To provide local school districts with information useful to their participation in the School Improvement Program.

The goals which relate to local-level information needs are:

To provide local school districts with information useful in making decisions about the allocation of resources and the design, implementation and evaluation of educational programs.

To provide local school districts with information useful in the assessment of students' progress.

The long-range goals of the Educational Assessment Program are to assess the status of New Jersey's educational system relative to all of the statewide goals which were determined by the "Our Schools" Project and adopted by the State Board of Education on April 12, 1972. The eleven "Outcome" goals pertain to individual student behaviors or attitudes desired as a result, at least in part, of formal educational experience. The nine "Process" goals describe how schools should function to provide meaningful education in New Jersey. The EAP will endeavor to provide information useful in the planning and evaluation of educational programs designed to bring the New Jersey educational system closer to these statewide goals.

The "Our Schools" goals are:

OUTCOME GOALS

To acquire basic skills in obtaining information, solving problems, thinking critically, and communicating effectively.

To acquire a stock of basic information concerning the principles of the physical, biological, and social sciences, the historical record of human achievements and failures, and current social issues.

To become an effective and responsible contributor to the decision-making process of the political and other institutions of the community, state, country and world.

To acquire the knowledge, skills, and understandings that permit him/her to play a satisfying and

responsible role as both producer and consumer.

To acquire the ability to form satisfying and responsible relationships with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.

To acquire the capacities for playing satisfying and responsible roles in family life.

To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.

To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the esthetic expressions of other people.

To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.

To develop an understanding of his/her own worth, abilities, potentialities, and limitations.

To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.

PROCESS GOALS

Insure that all instruction bears a meaningful relationship to the present or future needs and/or interests of students.

Insure that each student has significant opportunities, consistent with his/her age, for helping to determine the nature of his/her educational experiences.

Insure that specialized and individualized kinds of educational experiences are available for meeting the particular needs of every student.

Insure that teachers and students have significant opportunities for participating in the decisions affecting the operations of the schools they work in or attend.

Provide comprehensive guidance facilities and services of high quality for every student.

Seek to structure competition among students in ways that are less harmful than present practices are.

Insure that the resources available for education are used with maximum efficiency.

Insure that instructional, administrative and support staffs are of high quality in every respect.

Develop and utilize diverse forms of constructive cooperation with parents and community groups.

The Educational Assessment Program is intended to assist in the identification of local school districts which have conducted programs which have been especially successful and those which are particularly in need of assistance. In addition to providing local school districts with information relevant to the assessment of students' progress, the EAP will provide information to local districts that will be helpful in making decisions about the allocation of resources and the design, implementation and evaluation of educational programs.

TEST DEVELOPMENT

The Educational Assessment Program began in the fall of 1972 with the testing of fourth and twelfth grade students for achievement in the basic skills areas of reading and mathematics. In the future, testing will be conducted on grade levels four, seven and ten. Grade twelve students will be tested in three year intervals.

Lists of preliminary test specifications (Phase 1) were drawn up under the direction of the Division of Curriculum and Instruction in the State Department of Education. On June 5 and 6, 1972, meetings were conducted in each of the twenty-one counties of New Jersey by the county superintendents of schools for the purpose of obtaining a statewide preliminary consensus on the specifications for each of the four achievement tests. Each school district was invited to send four representatives, one for each of the four tests. Each person was given the appropriate list of test specifications, a machine-scorable answer sheet, and a sheet for comments. Respondents rated each topic according to the following scale:

A. Essential to be included on the test
B. Desirable to be included
C. Acceptable for inclusion
D. Inappropriate and should not be included on the test

Teachers were asked to rate topics on the basis of what is actually taught rather than on the basis of which topics they should be taught.

A total of 1,573 responses was received.

Rating forms and comment sheets were tabulated and reviewed. For each topic a mean rating was computed, and the topics for each test were rank ordered.

From this information Phase 2 specifications were prepared. The fourth grade reading test specifications and the fourth grade mathematics test specifications (Phase 2) were mailed to all third grade teachers in New Jersey; Phase 2 twelfth grade reading test specifications were mailed to all high school English teachers; and Phase 2 specifications for the twelfth grade mathematics test were mailed to all high school mathematics teachers. These 11,090 teachers were asked to rate the specifications and to add comments on separate comment sheets, using procedures similar to those used at the county meetings. This material was then returned to the Department. Approximately 7,950 responses were received.

The topics for each test were rank ordered on the basis of mean ratings. The responses of 6,375 teachers, or 57.48% of those to whom material was sent, were included in this tabulation. Some responses were excluded from the tabulation either because they were returned too late to be included or because directions for responding were not followed.

Four seven-member committees of examiners, one committee for each of the tests, were appointed to participate in the test development. Committee membership included teachers; administrators; representatives from the Black and Spanish-speaking communities; curriculum specialists; and, on the high school reading and mathematics committees, a high school student and a college student. Members were nominated by the Division of Curriculum and Instruction in the State Department of Education, and by key people in local districts and the State associations of mathematics teachers and reading teachers. These committees met on four occasions during July and August, 1972. These committees reviewed test specifications, reviewed and selected items (test questions) and agreed upon format and directions.

Items to meet specifications for each of the four tests (fourth grade reading and mathematics and twelfth grade reading and mathematics) either were drawn from files of existing items or were written.

Field testing of the first drafts of the tests was conducted on July 27, 1972. Because of the limitations of the nature and size of the samples, results of field testing were used only to check clarity of directions and format, to estimate overall difficulty level, and

to judge that groups taking the tests were reasonably comparable to groups on which items had been previously tested.

The drafts of the tests were reviewed by the Minority Groups Advisory Council. Members of this council are individuals who are familiar with tests and measurements practices or who are experienced in the education of minority groups.

The minority Groups Advisory Council met on August 14 to review the four tests, item by item, for cultural or linguistic bias. The Council recommended changes in some items on the fourth grade mathematics test, the fourth grade reading test, and the twelfth grade mathematics test. Most of the recommendations were rewording of items or redrawing of pictures to avoid unfavorable, stereotyped images of any minority group. The Council recommended major revisions in the twelfth grade reading test. It was the consensus of the group that paragraphs from contemporary sources should be added to this test to balance the traditional types of selections in the test.

To provide assistance in implementing recommendations, a subcommittee representative of the Minority Groups Advisory Council was formed. Since the consensus of the Council was that the subcommittee was representative of the Council as a whole, the subcommittee was authorized to act for the Council.

The subcommittee of the Minority Groups Advisory Council met with the twelfth grade reading test committee at Educational Testing Service. The group discussed each item on the twelfth grade reading test, and those items or words which were judged to be inappropriate were deleted. More appropriate words were substituted, and replacement items taken from contemporary literature were suggested.

TEST ADMINISTRATION

The tests were administered on November 14 and 15, 1972 to all fourth grade and all twelfth grade public school students in New Jersey. Practice tests were administered on November 9 in order to acquaint students with the format of the answer sheets and item types.

For schools which do not use grade designations, fourth grade students were defined as students who are in their fourth year of school beyond kindergarten, while twelfth grade students were defined as students who are in their twelfth year of school beyond kindergarten or who are in their last year of high school. Those students who are assigned to special education classes were excluded.

Tests were administered by classroom teachers or by other school personnel designated as test administrators by local district superintendents of schools. Training of test administrators in each district

was accomplished through building coordinators who were trained by district test coordinators. Administrative aspects of the program were carried out through local district and county test coordinators. These individuals were trained at one of three area training sessions conducted by the State Department of Education on October 11, 12, and 13, 1972.

TEST RESULTS AND THEIR USE

The following discussion deals with fourth grade mathematics test scores. The discussion is equally applicable to mathematics and reading scores for grades seven, ten and twelve.

Prior to the receipt of their test results, Local Education Agencies (LEA's) respond to a survey which lists the specifications around which the test was constructed. Examples of specifications for the fourth grade mathematics test are:

-- Addition of two 2-digit numbers involving carrying.
-- Addition of denominate numbers involving money.
-- Addition problems, involving no carrying.
-- Subtraction involving two 2-digit numbers with no borrowing.
-- The student can determine which of several sums is less than 60. Sums are in the form of 56 + 3, 58 + 2, etc.
-- Subtraction of two 2-digit numbers involving borrowing.
-- Subtraction involving two 3-digit numbers with borrowing.

Districts respond by indicating for each specification whether it is relevant to that district. Relevancy means that the district attempts to teach that specification at some time prior to the test period. The purpose of this survey is to allow the LEA's to increase the validity of the test by casting out items which measure specifications not relevant to that district. The procedure for discarding items is described below.

Because New Jersey exercises very little control over local curriculum decisions and because a uniform Statewide test is necessary, LEA's must have an opportunity to adjust their test results to compensate for differences between the LEA's curriculum and the Statewide consensus of curriculum objectives measured by the test. Although the discarding of items improves test validity, it must be remembered that LEA's are attempting many things not measured by this test or any other single test. The LEA's are encouraged to conduct additional assessments to measure progress toward these other objectives.

THREE QUESTIONS. There are three questions which should be asked by LEA's which use either these assessment instruments or any other tests. The questions are:

1. How are the children in the LEA progressing toward the objectives of the LEA?

2. How does the progress of the children in an LEA compare to the progress of children in similar LEA's?

3. What can the LEA do to improve its educational system?

Table 1 is a representation of the New Jersey Educational Assessment Program score report. This first report provides information relative to the first question.

The test results are reported item by item. The score unit is the number (and percent) of students tested in the LEA who answered the item correctly. No traditional norm-referenced scores are produced. Table 1 represents the results of a fictitious LEA called "Athens." The table format is identical to that of the school and classroom report. The report for an individual student indicates success or failure on each item. Table 1 provides information about seven items clustered under the heading "Problem Solving." The fictitious district Athens indicated that six of the seven specifications measured under "Problem Solving" were relevant to the LEA (Item 37 measures a specification which is not considered relevant). Approximately 83% of the 292 students tested answered Item 23 correctly. Provided with these results is a list which indicates the specification measured by Item 23 (and all other items). Athens now has some information about the mastery of that item. The State Department of Education does not determine mastery levels. Since local conditions determine the levels of mastery which can be expected, the LEA must decide whether 83% is "good" or "bad."

Item 37, which measures a specification this LEA said is not relevant, is reported in order to provide pre-instructional test data. Item 35, which measures a relevant specification, was achieved by approximately 56% of the students tested. The meanings of these results are not absolute; they are relative. Athens should inquire about the difficulty of the item, should consider when that specification was taught in relation to the test date, etc. The scores, then, are provided as guides, i.e., information to be used for decision-making, and are not to be considered extremely accurate, highly meaningful numbers. All test scores should be regarded in this way.

The mean of the percentages achieving each relevant item is called the Achievement Index Number (AIN) and is equal to 67.91. Athens can compute the AIN for each cluster and for all the relevant items taken together. By itself, the AIN has little meaning. It will

TABLE 1

NEW JERSEY EDUCATIONAL ASSESSMENT PROGRAM

DISTRICT RESULTS
NOVEMBER, 1972

LEVEL 4

SUBJECT MATH

DISTRICT Athens
NJSDE ID
NUMBER TESTED 292

ADDITION AND SUBTRACTION												
	ITEM	1	2	7	10	14	20	26	32	41	43	45	48
NUMBER CORRECT		281	285	188	263	251	235	245	239	206	188	160	185
PER CENT CORRECT		96.2	97.6	64.4	90.1	86.0	80.5	83.9	78.4	70.2	64.4	54.8	63.4
MULTIPLICATION AND DIVISION												
	ITEM	13	19	22	25	55						
NUMBER CORRECT		176	108	206	174	144						
PER CENT CORRECT		60.3	37.0	70.5	59.6	49.3						
MEASUREMENT AND GEOMETRY												
	ITEM	9	18	29	36	44	49	51	56			
NUMBER CORRECT		273	44	201	161	215	213	187	152			
PER CENT CORRECT		93.5	15.1	68.8	55.1	73.6	72.9	64.0	52.1			
PROBLEM SOLVING												
	ITEM	23	33	35	37	50	52	57				
NUMBER CORRECT		241	193	164	197	221	226	224				
PER CENT CORRECT		82.5	66.1	56.2	67.5	75.7	77.4	76.7				
NUMBER CONCEPTS												
	ITEM	3	8	16	27	30	31	38	39	42	46	60	
NUMBER CORRECT		218	259	264	214	179	169	228	181	238	147	173	
PER CENT CORRECT		74.7	88.7	90.4	73.3	61.3	57.9	78.1	62.0	81.5	50.3	59.2	
NUMBER SENTENCES												
	ITEM	4	11	12	15	21	24	28	47	53		
NUMBER CORRECT		253	265	50	204	266	249	128	165	146		
PER CENT CORRECT		86.6	90.8	17.1	69.9	91.1	85.3	43.8	56.5	50.0		
NUMERATION												
	ITEM	5	6	17	34	40	54	58	59			
NUMBER CORRECT		277	239	241	98	220	86	174	131			
PER CENT CORRECT		94.9	81.8	82.5	33.6	75.3	29.5	59.6	44.9			

TABLE 2

SUMMARY REPORT

MATHEMATICS LEVEL 4

MEDIAN INCOME $6,000-$7,000 (All Districts)

Cluster	Problem Solving						Number Concepts						Numeration					
Item No.	3	9	17	22	37	C	1	6	41	32	8	10	2	19	26	44	56	C
Per Cent	76	57	46	42	75	59	68	80	75	71	74	67	87	86	74	63	70	74

MEDIAN INCOME $6,000-$7,000 (Items Applicable to Districts)

Cluster	Problem Solving						Number Concepts						Numeration					
Item No.	3	9	17	22	37	C	1	6	41	32	8	10	2	19	26	44	56	C
Per Cent	78	64	59	47	73	61	72	81	76	76	80	72	88	84	78	55	73	78

take on more meaning when an attempt is made to answer Question 3.

Table 2 and other similar tables provide information which may be used to answer Question 2. This table permits Athens to compare its results to the results of LEA's serving communities with similar socio-economic status to Athens'. Many other comparison groups based on other correlates of achievement will be provided. The discussion is limited to this table.

Table 2 has two parts. The upper level provides information about students in all LEA's which have a median income between $6,000 and $7,000. The "per cent" is not an average, it is the actual percentage of students answering the item correctly. The lower level provides information on each item which reflects the percentage of students in LEA's which identified the item as relevant.

Table 1 and tables similar to Table 2 comprise the score report. The answer to Question 3 may be provided by Tables 3 and 4.

Table 3 provides information about the LEA. The variables listed here are considered modifiable, i.e., can be changed by the LEA. Table 4 represents a group of 47 districts which have a similar configuration of "hard-to-change" variables such as SES, educational level of parents, etc. These variables, in general, are related to achievement to a greater degree than are the modifiable variables. Tables 3 and 4 represent an attempt to examine the importance of the modifiable variables while controlling for the "hard-to-change" variables.

Assume Athens' Achievement Index Number for the entire test was 68. Athens would enter the frequency distribution in the interval 65-69. By comparing its own data (Table 3) to information about LEA's with a similar configuration of hard to change variables and which have AIN's approximately equal to Athens', Athens would discover that its pupil teacher ratio (31) was higher than the median of comparable districts (27); the average years of experience for Athens' teachers was 16 compared to 12 for the other comparable districts; Athens provided more special services personnel (1.2 per 100 pupils compared to .3 per 100 pupils); Athens' per-pupil expenditure was about the same as the other comparable LEA's. In reality, Table 4 would continue and provide information about all of the modifiable variables. The discussion here is limited to four variables.

LEA's are discouraged from making decisions about change on the basis of relationships between single variables and achievement. In reality, all variables interact with one another. Change in one results in change in the others. Since this program is intended to provide information for decision-makers by simulating reality, the simulation must include the interaction among all of the variables. By reading Table 4 horizontally Athens can see itself in relation to

other similar districts. By reading Table 4 vertically Athens can begin to see the importance of each variable in predicting achievement. To provide a more accurate statement of relationships between combinations of input variables and achievement, multiple regression equations will be provided.

The multiple regression equation based on information from these 47 LEA's will serve as a simulation model for Athens and the 46 other similar LEA's. Athens can "plug in" projected changes and compute projected achievement. Thus a proposed change can be "tested" prior to its implementation.

The description of the score report formats concentrated on LEA level reports. The same formats and procedures will be used to report results for school buildings, classrooms and individual students. Teachers can individualize the score reports based on the relevancy of test specifications to individual students.

In the Program's second year criterion-referenced evaluation, developed and implemented on the instructional (e.g., classroom) level, will begin. A teacher will then have information about individual students from the Statewide tests and information about the student from teacher-made tests. Each of these provides essential information and the assessment of student progress would be incomplete if only one system were employed.

We can no longer afford to make decisions affecting peoples lives simply by guessing the effect of those decisions--often after they have been implemented. We all must continually seek more and better information and the ability to use it properly.

TABLE 3

DISTRICT CORRELATES OF ACHIEVEMENT REPORT

District Athens

Correlate	Value
1. Class Size	33
2. Student Time (Day/Periods)	220 min/wk.
3. Race	W-25, N-50, P.R.-25
4. Enrollment	17,500
5. Teachers Experience (No. of years)	16
6. Age of School Building	26 years
7. Pupil/Teacher Ratio	31
8. Teachers' Salaries	$13,600
9. Teacher Experience (degree held)	A.B.-72, M.A.-25, D-3
10. Sex	F-86, M-14
11. Number Books in School Library	2200
12. Ethnic/Racial Composition of District	
13. Non-graded Classes/ Open Classrooms	0/0
14. Number of Graduates Entering Higher Education	62%
15. Number Instructional Rooms	63
16. Subject Offerings	
a. Number of Credits Required for All Pupils by Sub. Area	N.A.
b. Total Number Courses Offered in Subject Area	
17. Total Number Graduates	
18. Number of Admn. & Supervisors	60
19. Number of Teachers (total number)	565
20. Special Services Personnel (total number)	1.2/100
21. Non-professional Assistance (total number)	.3/100
22. Median Income	
23. Per Pupil Expenditure	$774

TABLE 4

Correlates of Achievement Reference Table (CART) No. 17

Level 4 Subject Mathematics

Achievement Index Number (AIN) From	To	Frequency	Cumulative Frequency	Median Pupil/ Teacher Ratio	Median Teacher Experience in District	Median No. Special Services Personnel Per 100 Pupil	Median Per Pupil Expenditure
95	100	1	47	23	17	.3	$1,200
90	94	3	46	26	18	1	948
85	89	2	43	22	18	.7	1,100
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
65	69	3	25	27	12	.3	786
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
0	4	1	1	29	3	.06	674

Bibliography

Evaluation in New Jersey Education: A Survey of Present Practices and Recommendations for Future Action. Trenton, New Jersey: State Department of Education. June 1970. 55 pp.

Our Schools--Second Conference on Educational Goals for New Jersey. A further Guide. Trenton, New Jersey: State Department of Education. June 20, 1970. 18 pp.

Our Schools--Second Conference on Educational Goals for New Jersey. A Report on the Outcomes. Trenton, New Jersey: State Department of Education. June 20, 1970. 7 pp.

Our Schools. Trenton, New Jersey: State Department of Education. No date. Six-page pamphlet.

"Statewide Educational Goals Proposed." New Jersey Education, Vol. 6: 3, March 1972.

"Public Opinion Poll Shows People Differ on Education in New Jersey." New Jersey Education. Vol. 6: 4-5, March 1972.

Our Schools--Your Opinions on Proposed Process Goals for New Jersey Schools. Trenton, New Jersey: State Department of Education. October 2, 1971. 12 pp.

Our Schools--Brief Descriptions of Process Goals. Trenton, New Jersey: State Department of Education. October 2, 1971. 55 pp.

The Governor's Conference on Educational Goals for New Jersey. Trenton, New Jersey: State Department of Education. October 2, 1971. 11 pp.

Governor's Message to New Jersey State Legislature, January 1972. Trenton, New Jersey: State Department of Education. 4 pp. (Calls for statewide testing for 1972-73 school year.)

Assembly Bill No. 822 (New Jersey State Legislature, March 13, 1972--pending). This bill was introduced to carry out the Governor's request in his message of January 1972 to the Legislature. 1 page.

A Resolution of the State Board of Education. Pertaining to the Establishment of Educational Goals for New Jersey. Trenton, New Jersey: State Board of Education. April 12, 1972. 3 pp.

Educational Assessment Program. Trenton, New Jersey: State Department of Education. May 10, 1972. 19 pp.

Educational Assessment Program--An Outline. Trenton, New Jersey: State Department of Education. June 20, 1972. 5 pp.

Educational Assessment in New Jersey: Questions and Answers. Trenton, New Jersey: State Department of Education. September 1, 1972. (Three versions. English and Spanish Editions for parents & students, 10 pp.; for teachers, 14 pp.)

Assembly Bill No. 1272. (Pending) By Assemblyman DeKORTE. (Concerns financing and disbursements of funds for public school districts.) Trenton, New Jersey: State Legislature. May 18, 1972 (Introduced). 21 pp.

Opinion Research Corporation. Highlights of a Survey on Educational Goals Conducted for New Jersey Department of Education. Princeton, New Jersey: Opinion Research Corporation. March 1972. 8 pp.

Senate Bill No. 2233. (Concerning teacher competency and certification.) Trenton, New Jersey: State Legislature. Approved May 20, 1971. 2 pp.

Performance Evaluation Project. Trenton, New Jersey: Department of Education and Rutgers College. November 1972. 14 pp.

*Performance-Based Teacher Certification: A Survey of the States. Trenton, New Jersey: Department of Education. December 1972. 15 pp.

Educational Goals for New Jersey (Tentative). Trenton, New Jersey: Department of Education. 1 p.

A Summary of the "Our Schools" Project. February 1970--February 1972. Trenton, New Jersey: Department of Education. March 1972. 41 pp.

Feldmesser, R. A. Educational Goal Indicators for New Jersey. Pfc-73-1. Princeton, New Jersey: Educational Testing Service, 1973. 66 pp.

A Resource Handbook for: Promoting Community Involvement in Educational Goal Determination. Trenton, New Jersey: State Department of Education. September 1972. 65 pp.

The "Consensus" Kit: A process Devised for Setting Educational Goals and Determining Consensus on those Goals Within a Group. Trenton, New Jersey: State Department of Education. 1971. 30 pp.

Opinion Research Corporation. Goals for Elementary and Secondary Public Schools in New Jersey. Princeton, New Jersey: Opinion Research Corporation, 1972.

Ascher, G. Individualized Instruction and Statewide Assessment: The New Jersey Educational Assessment Program. Trenton, New Jersey: State Department of Education. 1973. 22 pp.

Our Schools UPDATE, Vol. 1, No. 1, Trenton, New Jersey: State Department of Education, April 1973. 1 p.

New Jersey Statutes. 18A:4-24, Trenton, New Jersey.

Directions for the Word Recognition Section of the Fourth Grade Reading Test

Questions 1-4: (Clue words orally supplied by the examiner.)

1. "Find the word that begins with the same sound as *life*."
2. "Find the word that begins with the same sound as *ship*."
3. "Find the word that begins with the same sound as *bloom*."
4. "Find the word that begins with the same sound as *pain*."

Questions 5-8: (Clue words orally supplied by the examiner.)

5. "Find the word that ends with the same sound as *sat*."
6. "Find the word that ends with the same sound as *miss*."
7. "Find the word that ends with the same sound as *rang*."
8. "Find the word that ends with the same sound as *kind*."

Questions 9-19:

Students were directed to identify the word having the same sound as the underlined sound in the clue word.

Question 20-23:

Students were directed to find the rhyming pair of words in each group of words.

Questions 24-28:

The clue words in these items were orally supplied for the students by the examiner.

Questions 29-32:

Students were directed to denote the meaning of the underlined prefixes and suffixes.

Questions 33-38: (Directions read by examiner.)

33. "Choose the word that means more than one child."
34. "Choose the word that means more than one watch."
35. and 36. "Choose the compound word."
37. and 38. "Which two words mean the same as the underlined word."

From: New Jersey Educational Assessment Program Test Specifications and Items 1973-1974: Grade 4, N.J. State Department of Education, Office of Educational Assessment.

Test Specifications, 1973-74, Grade 4

Skill	Item No.
1. Word Recognition	
A. Phonetic Analysis	
Short vowels	11, 14
Long Vowels	9, 15
Vowel Digraphs	10, 13
Vowel Diphthongs	12, 16
Initial Consonants	1, 4
Variant sounds (c,g)	17, 18
Final Consonants	5, 6
Consonants Blends & Digraphs	2, 3, 7, 8
Silent letters	19
Rhyming words	20, 21, 22, 23
B. Structural Analysis	
Syllabication	
Identification of number of syllables	24
Correct syllabication	25, 26
Roots	27, 28
Prefixes	27, 30
Adjectival suffixes	31, 32
(Items 29-32: These items assess the student's knowledge of word meanings as well as their knowledge of prefixes and suffixes.)	
Plurals	
Regular	34
Irregular	33
Compound Words	35, 36
Contractions	37, 38
2. Reading Comprehension	
A. Word Recognition and Usage	
Context clues	49, 50, 51, 52, 53
Synonyms	39, 40, 41, 42, 43
	44, 45, 46, 47, 48
B. Organization of Ideas	
Identify Main Ideas	58, 65, 66
Understand Supporting Ideas	
Phrased exactly as in paragraph	54, 67, 68
Phrased differently	55, 60, 62, 69

3.	Understand Order and Reading Interpretation	
	Sequence of events	59, 63
	Determine character's motive	57, 61

(Items 57, 61: These items might also be interpreted as measures of inferential reasoning).

	Drawing inferences	56, 64
10	Study Skills	
	Guide Words in Dictionary	73, 74
	Alphabetical Sequence	70, 71, 72
	Table of Contents	77, 78
	Glossary	75, 76

NEW JERSEY READING TEST
SPECIFICATIONS
Grade 7

SKILL	ITEM NO.
I. Word Analysis (Word Study Skills)	
A. Phonetic Analysis	
Long vowels (i)	13
Short vowels(u)	14
Vowel digraphs (ea)	15
Vowel diphthongs (ou)	16
Single consonants (soft c & g)	17, 18
Consonant digraphs (ch, gh)	19, 20
Silent consonants (h, w)	21, 22
B. Structural Analysis	
Roots (identification of)	23, 24, 25
Prefixes (meaning)	26, 27, 28
Suffixes (meaning)	29, 30, 31
Compound words (identification)	32, 33
Syllabication	34, 35
Possessives	36, 37
Contractions	38, 39
Plurals	40, 41
C. Word Usage	
Synonyms	1, 2, 3
Multiple Meanings of a single word	4, 5, 6
Homonyms	7, 8, 9
Antonyms	10, 11, 12
II. Reading Comprehension	
A. Literal Comprehension	
Contextual clues	60, 61
Locating specific information (item phrased exactly as in paragraph)	44, 46, 47, 55, 58
B. Organization	
Sequence of events or ideas	65
Main idea	
Titles	43, 48
Unstated main idea	56

	ITEM NO.
Supporting detail (item phrased differently as in paragraph)	49, 51,* 52, 53, 57, 62
*Item 52 might also be interpreted as a measure of inferential reasoning.	
Cause and effect	54
C. Interpretation	
Drawing Inferences	42, 45,* 50
*Item 45 might also be interpreted as a measure of the student's ability to note specific information contained in the reading selection.	
III. Study Skills	
Following directions	67, 68, 69
Use of Maps	70, 71
Table of contents	76, 77
Index	78, 79
Glossary	83, 84
Dictionary Skills	
Use of guide words	74
Use of dictionary entries	72, 73
Alphabetization	75
Library Skills	82
Outlining	85, 86
Encyclopedia	80, 81

From: New Jersey Educational Assessment Program Test Specifications and Items 1973-1974: Grade 7, N.J. State Department of Education, Office of Educational Assessment.

NEW JERSEY READING TEST
SPECIFICATIONS
Grade 10

SKILL	ITEM NO.
I. Word Recognition	
A. Vocabulary	
Synonyms	9,10,11,12,13
Context Clues	1,2,3,4,5,6
Analogies	14,15,16,17
Antonyms	18,19,20,21 and 22
B. Structural Analysis	
Root words	23, 24
Affixes	25, 26
II. Reading Comprehension	
A. Interpret Reading Materials	
Identify Main Idea	
Titles	54
Unstated main idea	33, 39, 55
Understand supporting ideas	
(phrased exactly as in paragraph)	27,29,42,53 59
(phrased differently as in paragraph)	35, 41, 50, 60, 61
Identify sequence of events	31, 44, 56
See cause and effect relationships	28, 46, 47
Predict outcomes	32, 48*
*(Item 48 might also be interpreted as a measure of the student's ability to draw conclusions.)	
Draw conclusions	36, 37
Application of information in paragraph to other situations	52
Make inferences	34,38,40,43,58
Perceive tone of author's presentation	45, 51, 57
Recognize use of propaganda and arguments	30, 49
B. Other skills	
Interpret figurative language	7, 8

III. Location of Information

Dictionary	62, 63
Reference books	64
Card catalogs	65,66,67
Charts	68,69,70,71
Maps	72, 73, 74
Index	75, 76, 77
Table of contents	78, 79, 80

From: New Jersey Educational Assessment Program Test Specifications and Items, 1973-74, Grade 10. Office of Educational Assessment, New Jersey State Department of Education.

GRADE FOUR READING TEST RESULTS

GENERAL STATEMENT

The 1973-74 fourth grade reading test measured student's knowledge in the skill areas: Phonetic Analysis, Word Recognition and Usage, Organization of Ideas, Reading Interpretation and Study Skills.

The analysis of test results revealed that fourth grade students are most proficient in the areas of Phonetic Analysis, Word Recognition and Usage and least proficient in the area of Study Skills. Problems in Structural Analysis occur when students are required to denote root words when both prefix and suffix are added to the word, e.g., "indebted." Recognition of compound words, irregular plurals, comparative and superlative endings and certain contractions were also noted as problem structural analysis skills. In the area of Reading Interpretation, students experienced difficulties in locating main ideas, determining logical sequence and motives of story characters. They performed more capably in noting the supporting details contained in a passage.

The test results were analyzed according to fourth grade students' achievement in three general classifications: Community Type, Geographic Region of the State and the State as a whole. Except for a few instances, the Regional-Community Type and Northwest-Geographic Region achieved highest accuracy; however, differences from the scores achieved in the corresponding classifications were generally marginal. (Since consideration was given to item difficulty, but not necessarily to the variety of reading programs that are presently being used throughout local districts and the State, caution should be taken in respect to making definitive comparisons between classifications regarding achievement or lack of achievement).

CLUSTER INTERPRETATION

Phonetic Analysis

Performance in Phonetic Analysis was adequate in the entire cluster, when consideration was given to item difficulty, and variation and sequence of teaching decoding skills. Students demonstrated the most difficulty in identifying vowel sounds in words.

Grades 4, 7 and 10 Reading Test Results excerpted from State Report 1973-74 Educational Assessment Program, Office of Assessment, State Department of Education, Trenton, N. J.

Short vowels presented the greater problem followed by diphthongs and vowel digraphs. The depressed scores in these areas may be the result of problems non-English speaking students encounter in discriminating vowel sounds which are different from those that are contained in their native language. In the measures of ability to identify consonant sounds in words, students evidenced the most problems in differentiating final consonant sounds. In comparing results for Geographic Regions, the scores achieved by the students were close in range. Scores achieved in the various Community Types were relatively consistent except for the Urban classification.

Structural Analysis

Performance in the cluster Structural Analysis fluctuated with the degree of difficulty of the item. Students performed adequately when required to separate compound words into syllables and to locate one syllable root words. They had difficulty in both identifying compound words when these words were couched among others and locating root words when they were presented with a prefix and suffix added. Scores achieved for identifying the number of syllables in a word were adequate, but dropped when the task involved application of the vowel-consonant-vowel syllabication principle. Likewise, performance was adequate on the formation of a regular contraction, but was less adequate when the contraction was more complex.

Percentages of students' performance on Structural Analysis, when compared among the Geographic Regions, varied from one percentage point to five percentage points. The Northwest region scored highest, followed in the order of Southwest, Southeast and Northeast. The comparison of scores achieved within the Community Types ranged from three percentage points to eighteen percentage points.

Word Recognition and Usage

Students evidenced ability to make use of context clues and to identify synonyms. The troublesome areas that did occur were on items which included less frequently used terms, e.g., "trousers" and "sombrero." These words presented problems for students across the Geographic Regions and Community Types.

The range of percentages in the Geographic Regions extended from one percentage point to five percentage points. A wider range of scores existed in the Community Types, percentages extending from one to eighteen.

Organization of Ideas and Reading Interpretation

In the area of reading comprehension, highest achievement was attained when students were required to determine supporting ideas contained in a reading passage when the material related closely to everyday experience. Inadequacies in achievement were noted in students' ability to identify the main idea of reading passages, especially when the task required them to relate two or more sentences to arrive at the appropriate answer. Students also demonstrated limited ability to determine logical sequence of events, to interpret characters' motives and to draw specific types of inferences.

Differences between achievement, geographically, ranged from zero to four percentage points. Comparison of accuracy among Community Types indicated a spread of differences from five percentage points to sixteen percentage points.

Study Skills

Although students may have encountered difficulty in reading the directions that accompanied several of the items, performance in this cluster seems to point to weaknesses in the application of study skills. Many of these skills are taught in beginning reading programs; however, low test scores may indicate that students rarely apply these skills. Review of alphabetical sequence, table of contents, dictionary and the glossary should help strengthen students' knowledge and application of these study skills.

Submitted by: Shirl Burns
Ruth Kabot
Maria LaBoy
Agnarda Palsha
Francis Pinkowski
Judy Rock
Evelyn Slobodzian
Maria Stone
Mary Ann Wilmer

GRADE SEVEN READING TEST RESULTS

GENERAL STATEMENT

Results of the various skills tested: Phonetic Analysis, Word Usage, Literal Comprehension, Organization, Interpretation and Study Skills, seem to indicate that seventh grade students in New Jersey perform satisfactorily. Variations occurred in percentages on each item and were reflected in both the Geographic Region and the Community Types. The disparity of scores within the Community Type was

consistently larger. However, results also indicated that wherever one group performed inadequately on a particular test item, all other groups performed similarly. Conversely, where one group attained an outstanding percentage, other groups performed equally well.

Differences in scores were evident. Students performed well in identifying silent consonants, prefixes, compound words, contractions, use of plurals, multiple meanings of words, synonyms and map study skills. Levels of competency dropped considerably in the comprehension skills, such as those of context clues, locating specific information within a paragraph, sequential events, selection of titles, cause and effect relationship and drawing inferences.

Conclusions--etc. regarding differences in scores should be approached cautiously. Factors such as lower individual reading levels, language barriers, motivational factors, fatigue, passage readability and the ambiguity of particular test items were considered in the interpretation of these students' results. Statewide results indicated a slight disparity in scores reflecting the mastery of specific skills within each cluster. Community Type categorical scores show a higher performance percentage among the Suburban type as compared to Regional, Urban and Rural types. A definite pattern in performance was evident with the Northwest Geographic Region scoring consistently higher than other regions. Exceptions can be found however, in instances where low percentages were achieved in all Geographic Regions.

CLUSTER INTERPRETATION

Phonetic Analysis

Overall students' performance in the cluster of Phonetic Analysis was minimally satisfactory.

Performance was more than adequate in the area pertaining to silent consonants. In the measure of ability to identify long and short vowels and vowel diphthongs, students performed satisfactorily, except where the level of difficulty of an item was questionable.

Students experienced most difficulty in identifying vowel and consonant digraphs. The depressed scores in these skill areas could possibly be due to the construction of the test item. In comparing results for Geographic Regions, the scores achieved by the students were close in range (with the exception of Vocational). Scores attained in the various Community Types were relatively consistent, except for the Urban classification.

Structural Analysis

Overall achievement in Structural Analysis was considered satisfactory; however, student performance in this cluster fluctuated depending upon the skills involved. More than adequate performance was achieved by students in identifying compound words, plurals, and suffixes. Students scores were excellent when dealing with prefixes, contractions and identifying the number of syllables in a word. Structural analysis skills resulted in depressed scores when item construction was seemingly confusing and problem words were used. Although students in both Geographic Regions and Community Types achieved minimally satisfactory scores when they were required to locate root words and identify possessives, these skill areas should be emphasized in the instructional program.

Word Usage and Literal Comprehension

Students in all Geographic Regions and Community Types are proficient in Word Usage. They demonstrated ability to identify words with multiple meanings and to recognize homonyms.

When comparing test scores in Word Usage and Literal Comprehension, it appeared that students have knowledge of word meanings when words are presented in isolated situations, but have difficulty deriving meanings from words presented in contextual settings. Learned concepts of words are frequently altered when known words are presented in contextual situations, and thus puts realistic limits on the reader's level of comprehension. Particular attention should, therefore, be given to instructing students in making use of contextual clues to aide him in selecting the alternative meaning which a word carries in a sentence.

Organization of Ideas

Students evidenced greatest weakness in organizing ideas presented in reading passages. Inadequacies in achievement were primarily observed in the identification of main ideas and determination of supporting details. Noting the logical sequence of events contained in reading selections was also problematic. Scores attained on items in this cluster were consistent; when students achieved low scores or high scores on any item it occurred across all Geographic Regions and within Community Types. In most instances, the Urban areas achieved scores that were somewhat lower than the State scores.

Interpretation

An analysis of results for Interpretation revealed that students have difficulty in drawing inferences from written material. Inability to note the main idea of the passage contributed to the depression of these scores. This weakness was evidenced by students

throughout the State.

Study Skills

Generally students performed inadequately in the cluster Study Skills. When consideration was given to item structure and methods of measuring particular skills, some scores were interpreted as minimally satisfactory i.e. (using table of contents, using the glossary and following directions).

Pronounced weaknesses were observed in the area of dictionary skills, library skills, alphabetization, outlining and use of encyclopedia.

Low scores were achieved in the above areas by students in all Geographic Regions and all Community Types.

Submitted by: Donald Beineman, Katherine Brown, Roscoe Brown, Donald Clark, Marie Adair, Ronald Lewis, Evelyn Matthews, Joyce Matthews, Adele Stern, James Van Zoren

GRADE TEN READING TEST RESULTS

GENERAL STATEMENT

The 1973-74 Educational Assessment Program 10th Grade Reading Test examined reading skills in the New Jersey public schools in four main subjects areas (clusters): Vocabulary, Structural Analysis, Interpreting Reading Materials, and Location of Information. Each of these clusters if further sub-divided to test individual skills.

Students in Regional and Rural communities performed consistently higher than students in Urban and Vocational Communities. In the case of the vocational schools, the vocational curriculum precludes devoting as much time to academic skills as other Community Types are able to devote. This may account for the lower scores. However, the reasons for the lower scores of the Urban and Vocational Students should be investigated.

These tests indicate Statewide that students need more training in vocabulary development, critical reading skills and location of information. In the cluster measuring their knowledge of vocabulary, students performed well analyzing the meaning of words in sentence

context, but performed lowest in the sub-clusters testing words in isolation. In the cluster Interpreting Reading Materials, students had the greatest difficulty in the sub-cluster Making Inferences, indicating that critical reading skills, under this category particularly, need more attention.

Under Location of Information, students' scores showed a weakness in using the library card catalogue and reference card. The need for more training in these library skills is indicated, since the pursuit of any form of study or research calls upon the students' skills in this area.

While the test scores of the Community Types varied in some instances by more than twenty per cent, the test scores of the four Geographic Regions varied in almost all cases by less than five per cent. Generally, the Northwest scores were higher than those of the other regions and the Southwest scores lower.

CLUSTER INTERPRETATION

Vocabulary

Of the twenty items in this cluster, only three elicited scores that were unsatisfactory. General performance for the Vocabulary cluster was good. However, students' scores were consistently higher when the items tested vocabulary in context than when vocabulary was tested in isolation (synonyms and antonyms). In fact, the antonym sub-cluster produced the lowest scores of all the vocabulary sub-clusters, with a statewide spread of scores from fifty three to seventy seven (per cent correct). In the synonym sub-cluster, the students' performance was only a little better, with a variation in score from fifty per cent to ninety two per cent.

In contrast to their difficulties with antonyms and synonyms, students performed much better analyzing the meaning of words in context. Under the sub-cluster Context Clues, statewide scores varied from seventy six per cent to ninety two per cent with a higher percentage in general for individual items.

In the sub-cluster on analogies, the scores were consistently high with a Statewide variation in percent correct from seventy eight to ninety two. These scores seem to indicate that, by and large, students have a good grasp of word relationships.

The difficulties of students experienced in identifying synonyms and antonyms in contrast to the strength they demonstrated in analogies and context clues indicate that more attention needs to be focused on vocabulary development. The encouragement of greater usage of the dictionary, with a sharper attention paid to the meanings and nuances of individual words, might prove more beneficial

than just vocabulary drill or rote memorization.

Structural Analysis

Statewide results were better for affixes than root words in this cluster. Vocational and Urban scores were closer to the Statewide averages in this cluster than in the other clusters.

The test results seem to indicate that a greater stress is needed on the understanding of the concept of root words and on their identification.

Interpret Reading Materials

This cluster comprised the largest section of the tenth grade reading test. Of a total of thirty-seven items, the students performed unsatisfactorily on three (eight per cent), minimally satisfactorily for eight (twenty-two per cent) and satisfactorily or better on twenty-six (seventy per cent) of the items. Therefore, the overall performance of the students in interpreting reading materials was adequate.

In the sub-cluster Making Inferences the scores were the lowest in all of the cluster, indicating that critical reading skills, under this category particularly, need more attention.

Students did well identifying the main idea of a reading passage and in understanding supporting ideas when the item was phrased similarly to the paragraph (the items elicited details stated directly in the paragraph). Students experienced more difficulty when the phrasing in the item differed from that of the paragraph, again indicates the need for more stress in developing their critical reading skills

The test results for this cluster indicated a relationship between the timeliness and interest level of the reading passages and the scores of the students. That is, the more stimulating and timely the passages, the higher the scores. This correlation is particularly noticeable in the scores of the Vocational students.

The Urban scores were higher than expected for this cluster.

Location of Information

Overall scores were judged adequate. The performance of the students in the sub-cluster card catalogue indicate the need for more adequate training in understanding how to use the card catalogue and the library reference cards.

The Vocational and Urban scores were closer to the State average for this cluster.

Submitted by: E. Julia Basco
Warren Cummings
Joseph Flannery
Norman Lemberg
Gibson LeRoy
Patricia Mull
Vida Nogi
Sylvia Oettle
Richard Sheperd

THE BIG PICTURE

HOW MANY PUPILS AND TEACHERS ARE THERE IN NEW JERSEY?

N.J. School Population 1973-1974

Administration and Supervision	6,502
Classroom	79,156
Special Services	9,065
TOTAL CERTIFIED	94,723
NOT CERTIFIED	35,855

Pupils

Elementary K-6 and Middle Elementary K-8	953,284
Secondary	528,321
TOTAL	1,481,605

Estimate of non-public pupils

K-8	174,158
High School	65,726
TOTAL	239,884

Number of Pupils in Tested Grades

4th grade	113,542
7th grade	116,807
10th grade	116,904

Excerpts from

GUIDELINES FOR THE INTERPRETATION

OF THE NEW JERSEY EDUCATIONAL

ASSESSMENT PROGRAM RESULTS

1973-74

INTERPRETATION

A. Materials received at the District Level - Each district should receive the following materials:

- Summary reports for District, Building, and Classroom.
- Summary reports (both "relevant" and "totals") for community-type, county, geographic region, and State.
- Individual student data by classroom.
- Summaries of General Information item responses.
- Interpretive materials to allow an appropriate set for each building.
- Camera ready copy of interpretation worksheets.

1. Summary Tables -

a. District, Building, and Classroom reports -

Each district will receive their test results for the district level, the building level(s), and the classroom level(s). You may also note that items measuring the same specification are grouped to allow easier interpretation.

b. Comparison Groups -

Each district will also receive color-coded summaries for community-type, county, region, and State results.

Office of Educational Assessment, State Department of Education, Trenton, N. J.

2. Individual Student Data -

Responses to the general information items are presented as well as the results on each specification within clusters. By scanning the list by name, the teachers may note the individual strengths and weaknesses of each child. By scanning the lists vertically, the teacher may note the strengths and weaknesses of the class as a whole.

3. General Information Summaries

For each grade level tested, a set of general information items were developed and asked of each student. The Individual responses to the general information items are presented on the individual student results tables.

4. Test Specifications and Items Booklets -

With the district results, each district should receive a copy of the test specification booklets for each grade level tested within the district. Each district will also be provided enough copies of these booklets, through the county office, to provide each school with at least one copy at the appropriate grade levels.

5. Interpretation Worksheets -

A complete set of interpretation worksheets will be distributed to each district from the county office. These worksheets may serve as camera-ready copy for any reproduction deemed desirable by the district.

B. Phase I Interpretation

The Phase I interpretation is designed to address the question:

To what degree is the district (school, classroom and student) achieving its stated objectives?

For all tests (reading and mathematics) at all three levels (fourth, seventh, and tenth), the format of the results and the procedures for interpretation are the same.

The item number and specifications are pre-printed for district use. (You may wish to refer to the Test Sepcifications and Items booklets). The district needs to enter only the relevant notation and the percent correct observed from the district results.

The "relevant" notation, printed with your test results, indicates whether or not the item was considered to be relevant or not to your district (Y=Yes, N=No). You may recall that your district was asked to respond individually on each test specification as to whether the specification was taught by the end of the previous grade level. This survey was used to prepare the Phase II relevant comparison groupings as well as to note your district's unique curriculum specifications.

For the Phase I interpretation, using your knowledge of your curriculum and your instructional program, examining the actual test item and the item specification, and considering any information gathered by other formal and informal means, make a judgment as to whether the item result:

(1) Is satisfactory and requires no change in curriculum and instruction.

(2) Indicates some need because the result is less than your expectation. If so, determine in concert with appropriate staff, possible reasons for observed needs and consider means for addressing these needs.

Types of considerations in a Phase I interpretation would include, but would not be limited to, the following:

(1) Is the specification relevant to the curriculum?

(2) To what degree does the specification receive emphasis and reinforcement in the curriculum?

(3) How difficult is the specification concept and the item?

(4) How recently have the students been exposed to the specifications?

Under the column headed "Item Specification Comments" the interpreters would enter whether the specification represents a need or a strength and would suggest appropriate alleviation of needs. For the Phase I column titled "Cluster Comments," the interpreters would summarize the item specifications comments across each cluster.

(Note: See Table I for the format for Phase I Interpretation.)

TABLE I

Test Level _____
Subject _____
Cluster Title _____
County _____

Item Number	Specification	Relevant	Percent Correct Observed	PHASE I INTERPRETATION	
				Item Specification Comments	Cluster Comments

TABLE II

Test Level _____
Subject _____
Cluster Title _____
County _____

Item Number		Community Type	County	Geographic Region	State	PHASE II INTERPRETATION	
						Item Specification Comments	Cluster Comments
32	T						
	R						

C. Phase II Interpretation

The Phase II interpretation is addressed to the question:

Is the District's Achievement comparable to that of representative groups?

Each district receives results for their community type, county, geographic region, and state for both "relevant" and "totals."

The Relevant summaries are composed of the results of districts who considered the item relevant to their curriculum

The Totals summaries present the results of all students in the grouping regardless of whether the item was considered relevant or not.

The districts will be provided with camera ready copy for Phase II sections of the worksheets which will have the state results, appropriate geographic region results, and appropriate county results printed.

The district is required to enter their community-type results. If a district desires to use an additional community-type comparison, they should address a letter to the Office of Assessment stating reasons at the beginning of interpretation.

The Phase II interpretation example may be noted in Table II.

(A) Using the worksheet, compare the district's results on each item with each of the comparison group results.

(B) For each item, note the items which indicate greater achievement than the representative groups and those which indicate lesser achievement than the representative groups.

(C) Note the items that you determined as needs in Phase I and compare the district percentages to those of the representative groups. You may find that an item with a low district percentage compared to the other items in that cluster may represent a depression within the representative groups as well for that item compared to that cluster. This may be of use in revising the list of needs developed in your attempt to answer Phase I. The same procedure may be followed for items which are higher than the cluster in the district report. For

the Phase II column titled "Cluster Comments," the interpreters would summarize the item specification comments across each cluster.

PRODUCTS

The procedure described for interpretation will result in reports which must be submitted to the Commissioner of Education for approval and, when approved, released to the public along with the original data received by the local district. A checklist of products and accompanying information to be included in this report follows:

A. District Level Report

Each district must prepare and submit a report for each test administered in the district to include the following:

PHASE I

1. A list of all curriculum specifications measured by the test.

2. Item numbers associated with each specification.

3. Notation of which curriculum specifications were considered relevant and which were not relevant (these must be presented by specification and item number and grouped by cluster headings.)

4. The percent correct observed for each item on the district level presented by specification and item number and grouped under broad cluster headings.

5. Item specification comments:

   a. Notation of the specifications which have been achieved to the district's satisfaction (strengths).

   b. Notation of the specifications which have not been achieved to the district's satisfaction (needs).

6. Cluster Comments:

   a. Summaries of specification strengths and needs.

PHASE II

7. Each report must provide: the summary results for the State, the appropriate geographic region, the appropriate community type, and the appropriate county for both "relevant" summa-

ries and "Totals" summaries. These results must be grouped by cluster and presented with the district results and the test specifications.

8. Strengths and needs statements for each item specification based on Phase I district results and comparison groups results must also be provided.

9. Summary statements, by cluster, of specification strengths and needs.

10. Names and positions of persons assisting in the interpretation. NJAC 6:39-1 et seq. requires Superintendent involvement on the district level, principal involvement on the building level, and teacher involvement on the classroom level.

B. Building Level Reports

A building level report for each building in which testing took place is required to be submitted to the state for approval. These reports must satisfy the Phase I requirements noted in items 1-6 above. In addition, item 10 names and positions of persons, must also be satisfied.

The requirements above apply to fourth grade reading, fourth grade mathematics, seventh grade reading, seventh grade mathematics, and tenth grade reading. For the tenth grade mathematics test, however, five separate reports are returned to the district based on the math background of the student. An interpretation is required for each of the four levels where the number of students tested is equal to or greater than twenty. The fifth report, the summary of the proceeding four reports, does not require district or school level interpretation.

Reports using averages or other statistical measures of central tendency will not be approved.

C. Report Options

1. Reproduction of test items which measure relevant specifications.

2. Possible reasons for observed strengths.

3. Possible reasons for observed needs.

4. Possible procedures for the alleviation of needs.

5. Narrative Summaries.

D. Classroom Level Reports

1. Classroom level reports require the completion of items 1-6 under District level report.

2. Classroom level reports are not to be submitted to the State Department of Education for approval except in cases where the classroom also represents the building and/or district level report. In these cases, the classroom reports should then carry the identification of Building Report or District report.

3. Classroom level reports are designed to provide feedback to the teacher and are generally used as data in developing the building level reports.

To Perform and To Account by Roger T. Lennon

Journal of Research and Development in Education, Vol. 5, No. 1, Fall 1971.

Pages 50-52 were removed due to copyright restrictions.

IT'S EASY TO MAKE LOCAL NORMS

by

Edward Fry
Director, Reading Center
Rutgers University

Teachers sometimes think that you have to have a computer or at least a statistician to make local norms. They are wrong. Almost anyone can quickly learn to make local norms, and you do not need any statistics, computers or even adding machines. If you can count and have a piece of graph paper, you have all you need.

There are many reasons for making local norms. Standardized tests usually give only national norms and might give only part of the information you want in your school or district. Informal tests written by teachers, criterion referenced tests, and all types of skills tests to measure objectives might not have any norms at all. It is useful to have local norms to see how an individual student ranks in comparison to his peers. It might be important for the teacher, the supervisor, or the child to know that he moved from the fourteenth percentile to the sixty-fifth percentile after instruction.

A percentile tells how a student compares with a standardization or norm group. A percentage is almost the same as a raw score; it tells only what percent of the total possible items a student got right.

The trouble with raw scores, which are the kind of scores on most teacher-made tests and criterion referenced tests, is that they are determined by the test author alone. If the author says that 90 percent is "passing" it is often based on nothing but the author's subjective judgment. What if the author is wrong, and it is almost impossible for any child to get 90 percent. Or if anybody, without even studying, can get 90 percent of the raw score? Normed tests attempt to eliminate this difficulty by comparing the student with his peers. In many respects, it can be much fairer to the student than an authority's judgment.

To be really fair you must ask, "With which peers do I compare this student?" Most published standardized tests attempt to use children all across the United States from a similar aged group. This is one useful comparison, but it has some real pitfalls. It includes regional differences, class differences, ethnic differences, home influence differences, curriculum differences, differences in school philosophy and emphasis, but these differences are often submerged by

From: Journal of Reading, Vol. 18, No. 3, December 1974, pp. 241-246.

53
PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
International Reading Association
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

the "majority." If phonics is not stressed or taught much in a local school, why should students be penalized by getting low phonics scores in national norms? It is more useful to know, that Juan is in the top 20 percent of his class taking a reading test in English than that he is in the bottom half nationally.

If you are convinced that local norms might be useful in measuring achievement or progress, here is how you make them. You can use raw scores (the number of questions answered correctly) on either a formal published test or a nonstandardized test such as one a teacher or curriculum committee made up. It will not make any difference if the test is a little too hard or a little too easy for the group -- local norms adjust for that. You must get a range of scores, however, so that some students do better than others, and so there are not extremes at the top or bottom of the test score range.

Start with an easy example. We will assume that you have 100 students in your fifth grade, and you want local norms for the fifth grade on a test. You administer a reading test which has fifty items; the lowest student gets fourteen right and the best student gets forty-nine right, so the range of scores goes from fourteen to forty-nine.

1. Sort the answer sheets so they are in order of raw scores: fourteen on the bottom and forty-nine on top. In the middle of the range, a number of students might have received the same raw score of thirty-six.

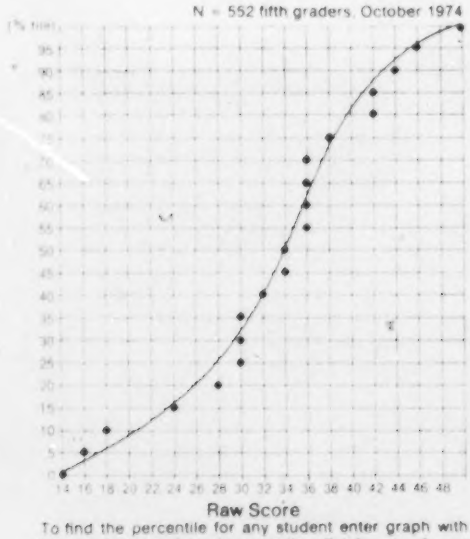
2. Label a sheet of graph paper so that along the vertical axis the lines are numbered from 0 to 100.

3. Label the columns with the range of raw scores, in this case from fourteen through forty-nine.

4. To plot each score on the graph, the first score would be represented by a dot at the intersection of line one and column fourteen; the second paper (which had a score of sixteen) would be at the intersection of line two and column sixteen; the third paper is plotted on line three, and so on. If you are in a hurry you can merely plot every fifth paper and not lose much accuracy. What you should end up with is a group of dots which more or less form an S-shaped curve on the graph paper. (See figure)

Local Norm Graph Converting Raw Score to Percentile

XYZ Reading Test
Form A



To find the percentile for any student enter graph with his raw score and read percentile off left axis. A raw score of 34 places student at 50th percentile in comparison with the norm group of 5th graders.

5. Draw a curved line, which mathematicians call "smoothing the curve," so that it seems to go through the average of the dots. It will actually go through some dots, but just above or below others. Do not bend your curve so that it goes through every dot. It must be basically smooth without bumps and sags. Usually it will have a slight "S" shape. This, incidentally, is a very sound mathematical procedure; doing it graphically is easy, while doing it by formula is very hard. After you have penciled in your curve, you might look at it from a distance, make a few adjustments, get rid of irregularities, and finally ink in the finished curve.

6. Your local norms are ready for use after labeling the vertical column (1 to 100) "Percentiles."

Use of Nomograph

To find a percentile with a raw score of forty, you go to the column above the raw score of forty, see where it hits the curve, then read the percentile across from that point, which is eighty-one according to the chart.

This curve on the graph paper is called a nomograph, and for most practical purposes, it is just as accurate as a table of numbers which convert raw scores into percentiles which are found in the typical test manual.

The trick of this whole thing is not drawing the curve, but in selecting the norm group. It should be representative and as big as possible of that representative population. If, for example, you have 552 fifth graders, you should put all the answer sheets in a pile, shuffle them as well as possible, then randomly draw out 100 papers to plot. You can, of course, make local norms with less than 100. You will not get quite as stable a norm base, but in many instances it might be much better than national norms for your purposes to see how your own individual children do.

Local norms can be made each year or the same graph can be used for several years if you continue to use the same test. They can greatly aid the reading teacher, administrators, or classroom teachers in using and interpreting reading test scores for published standardized tests, criterion referenced tests, or locally made tests.

Excerpts from

USING GROUP READING ACHIEVEMENT TESTS TO IDENTIFY PROBLEM READERS

Roy A. Kress

Dan was fifteen years of age and in the ninth grade when the author first met him. He was a handsome lad, of normal stature for his age, and had been brought to the university classroom to participate in a demonstration of certain tests frequently used in the analysis of reading disabilities. Among the facts reported about the history of his difficulty was the information that it had "first been discovered while he was in fifth grade" although the school record showed that he had repeated both first and fourth grades. Currently he was receiving forty-five minutes of "remedial" instruction in reading each week and his ninth grade English teacher had adjusted her daily classroom instructional program in reading for Dan to his "fifth grade reading level." Her decision to work with him at this level had been based upon the . . . array of test scores obtained from the Iowa Basic Skills battery administered the preceding March when Dan was in eighth grade.

. . . . . .

A California Mental Maturity Test, Short Form, had been administered within the month (October) and reflected a total I.Q. score of 77, with little difference between his language and non-language sub-test scores. Such a score for a fifteen-year-old would yield an estimated mental age of approximately eleven years and indicate an expected academic achievement level of Grade 5½. Dan was, to all practical appearances, a very slow learner, had been throughout his school career, and was functioning in reading achievement at about a level which could be expected of him--fifth grade. The teacher thought so! The standardized reading achievement tests indicated so! And the mental test results confirmed these opinions!

In contrast to Dan's performance on achievement and mental ability tests designed for measurement of the academic group in which he had been placed, individual tests of these areas provided additional

From: Problem Areas in Reading---Some Observations and Recommendations. Rhode Island College Reading Conference Proceedings, 1965. Coleman Morrison, Editor.

information. On the Wechsler Intelligence Scale for Children he obtained a Full Scale I.Q. score of 83 with a performance score of 92 and a Verbal score of 77. The psychologist who administered the test reported these scores to be sub-minimal estimates and concluded: "Dan's potential level of academic functioning, given adequate help with his reading problem, is certainly within the average range and may even be at the bright normal level."

On an Individual Word Recognition Test Dan was unable to recognize more than 55% of a randomly selected list of the words normally introduced in a typical First Reader, even though he was given all the time he wished to apply any word analysis skills at his command. None of these words was in his sight vocabulary. When an Informal Reading Inventory was administered, Dan's performance indicated complete frustration at the Pre-Primer levels of materials, although his listening comprehension was found to be satisfactory at the so-called Eighth Reader level. . . .

. . . . . .

This description of Dan's performance serves to point up the extent to which scores from group tests of reading achievement have been improperly utilized for making decisions about the instructional needs of individual children. Dan's teacher, acting on the notion that the achievement test last administered accurately reflected the level of his instructional need, was making a genuine effort to meet this need at the "proper" level. It is understandable that both she and Dan were discouraged, and also that she had developed the feeling that grouping for instruction was fine in the elementary school, but it wouldn't work at the secondary levels."

Reading specialists could enumerate many cases like Dan. The literature on research in reading cites numerous instances of the over-prediction of proper working levels for children resulting from the adherence to a set of scores from group tests of reading achievement. Some researchers have recommended the automatic subtraction of two grade levels from a child's grade score (secondary levels) to determine his proper instructional level in reading. Others have suggested that such tests place children at their frustration reading levels. However, Sister Davis (2), in evaluating the relative effectiveness of standardized reading achievement tests in determining instructional reading levels for fourth grade children, found a wide range of discrepancy in scores when compared with individually administered informal reading results. Although her correlations were all positive, they were low and the mean differences between standardized and informal measures ranged from -3.2 to +2.7 grade levels. She concluded that standardized tests of reading are not accurate in defining any level of reading achievement. She considered the grade scores obtained from such tests as likely indications of a "pre-frustration range" for most students, but found enough exceptions to

warn against the application of any rule-of-thumb formula for determining instructional levels from standardized group tests.

Of what value, then, are group standardized reading tests? Should they be discarded completely, or can they be utilized in identifying the problem reader? If the latter be the case, what information will they yield and what cautions should be observed?

Group measures of reading achievement are primarily designed to yield information about the performance of a group of individuals on the test selected. Properly administered and scored, they provide a legitimate basis for group comparisons in reading achievement. In some instances, the groups may differ because the individuals within the groups are not the same people (Class A vs. Class B). In another case, the differences may be one only of an interval of time (Class A, September vs. Class A, June). When the tests are administered and scored in the manner in which they were standardized, the group tested may be compared with the national, regional or local sample used in the standardization. It would also be legitimate to compare the results of two different tests for the same group of children. . . .

. . . . . .

It seems important to point out that when we depart from the use of group measures for the comparisons of the characteristics of groups to the use of individual scores obtained in such tests we tread on very perilous ground. There are many factors other than ability to achieve at a particular level or in a particular situation which may influence an individual's performance on any given group test of reading achievement. The accuracy with which these factors may be identified will be increased only when individually administered reading achievement tests are employed.

. . . . . .

SUMMARY

Group standardized reading tests are of limited value to the classroom teacher or the reading specialist for identifying the child who has a reading problem. Although they can be used effectively as pre-post indications of instructional gain, or in the measurement of the mean reading achievement level of a group of children, both validity and reliability are lost when the specific scores of individual children are considered. At best, they may be used as one of a series of screening measures in identifying the children in a group who appear to be candidates for more careful and detailed study to determine whether a reading problem actually exists. There is no evidence to indicate that group standardized reading tests can be

employed successfully to identify the proper instructional reading level for children, either in the regular classroom or in corrective and remedial reading programs.

REFERENCES

1. Bond, G. L. and Fay, L. C. A comparison of the performance of good and poor readers on the individual items of the Stanford-Binet Scale, Form L and M. Journal of Educational Research, 1950, 43, 475-479.

2. Davis, Sr. M. Catherine, RSM. The relative effectiveness of certain evaluative criteria for determining reading levels. Doctoral Dissertation, Philadelphia: Temple University, 1964.

3. Englehart, M. D. Improving classroom testing. What research says to the teacher, Bulletin No. 31. Washington, D.C.: National Education Association, 1964.

4. Lyman, H. B. Test scores and what they mean. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1963.

5. McLaughlin, K. F. Interpretation of test results. U. S. Department of Health, Education, and Welfare Bulletin EO-25038, No. 7. Washington, D.C.: Government Printing Office, 1964.

A Word to Teachers on Statewide Testing in New Jersey

Margaret Smith-Burke
Department of Educational Psychology
New York University

Several questions have been raised by classroom teachers and administrators about the interpretation of the statewide test results in reading. How can teachers use the results? What effect might test results have on reading curricula in the state? Can the tests be used diagnostically? Before these questions can be answered, it is important for teachers to understand their own thoughts and assumptions about the reading process.

Evaluation of the statewide reading tests and interpretation of their results will be much easier, if the teacher starts with a model of the reading process in mind. The following questions should help to create this type of framework.

1. What is the purpose of reading?
2. What general skill areas are included in your concept of the reading process? Visual and auditory discrimination? Language development? Concept development? Word recognition? Comprehension? Study skills?
3. How are these areas and their subskills related to one another? For example, should a child learn all consonant sound symbol correspondences before learning vowel correspondences? How does the child's dialect and oral language development affect the teaching of sound symbol correspondences? Must all sound symbol correspondences be mastered before comprehension is taught?

4. Specifically, what skills are taught in your reading program, how and when?

Armed with a concept of the reading process and a summary of how, when and what skill areas and subskills are covered in her reading program, the teacher is ready to critically assess the statewide tests and interpret her class results. The concept or model of reading helps her weigh the importance of each skill relative to other skills.

The first task is to compare the content of the test to the content of the reading program. How similar are they? If there is little similarity between the content of the program and the test, the teacher may conclude that the test is inappropriate for her group and has little relevance. At this point, a questioning teacher might wish to explore why there is so little similarity and whether her colleagues share her perceptions of the test.

Have other members of the faculty also arrived at this conclusion?

If a majority of the faculty feel the content of the test does not emphasize the skills being taught, there are two possibilities:

1. It might be worth considering modifying the reading curriculum. Have any critical skill areas been omitted from the curriculum?

2. After discussion of the mismatch between test and reading program content, if the faculty still feels that the test incorrectly emphasizes certain skills and omits others that are currently being taught, notification of the assessment program personnel might be considered to suggest how the test could be improved. Since the types of items included in the statewide reading tests are developed from a list of skills ranked in order of teaching priority by New Jersey teachers, the faculty members should report in detail the skills actually taught in their schools.

If there is high or at least partial correspondence between the test content and the reading program, the teacher can interpret the results of the test. If all of the children in the class perform well on the test, the teacher can breathe

a sigh of relief and wait to face the problems of interpretation until next year. However, if a substantial percentage of the class has difficulty with certain items, the teacher must be able to interpret these results and try to hypothesize why this has happened. There are three helpful caveats to keep in mind when undertaking the interpretation task.

First, in order to test any skill, there must be an adequate sample of questions measuring that skill. If someone is testing a person's ability to drive a car by asking him to parallel park one time, how reliable is the measure based on that one attempt? In other words, there must be enough items testing a particular skill to insure that the children really know the skill and are not getting the item correct for a variety of other reasons. On the fourth grade test on reading there are only two items which test two different short vowel sounds. Does this really measure knowledge of short vowels? On the seventh grade test there is only one item on understanding cause and effect. Since there are so few examples to measure any given skill, it would be inadvisable to use these results for diagnostic purposes, particularly for individuals. At best, the results might provide a general idea of the areas in which further diagnosis might be pursued, if the teacher feels the need of additional information.

Another question to keep in mind when interpreting the results is how the skill has been measured. What type of testing task was used? Does this task correspond to the types of activities which have been used in the classroom? For example, in teaching the concept of main idea a teacher asked the children to find the sentence in the passage that best expressed the main idea, or to paraphrase the main idea in their own words. On the test the items called for the choice of the best title of a passage. In this case there was a mismatch between classroom behaviors which demonstrated the ability to understand the main idea and the behavior measured on the test. If this is a problem, the teacher may wish: 1) to disregard those particular questions, and/or 2) to modify her teaching method to increase the transfer of a given skill to a variety of situations.

Finally, the relationships of skill areas and their subskills implied in the teacher's model of reading are important in interpreting results. The extent to which subskills can be isolated should be considered. If the subskill

can be tested in an isolated situation, what does knowledge of this isolated skill mean in relation to the complex process of reading in context? The importance attributed to a particular subskill and/or area of skills depends on the teacher's model of reading. For example, a teacher whose model of reading stresses meaning and whose class performs well on the comprehension questions, may choose to give less weight to the phonics results since her model implies that adequate responses to the comprehension questions reflect integration of phonic skills with other reading skills. This might well be the case on the seventh grade statewide reading test. Another example is a teacher who has a number of bilingual Spanish-English speaking children in her class. She would expect these children to have certain problems on isolated short vowel sound symbol correspondences which they might not have when encountering the same correspondences in meaningful context.

Returning to the questions which have been posed about the use of statewide testing scores by teachers, one finds that two have already been answered. Even though the teacher should not use the results diagnostically for individuals, she can assess on a class basis how successful she has been in teaching reading in general. However, in order for her to interpret the results of the tests intelligently, she must be aware of her own concept or model of the reading process and how the skill areas and their subskills interrelate. From the results she will have a general idea of the skill areas which are in need of further diagnosis.

It is premature to judge how statewide reading tests will affect reading curricula in the state of New Jersey. One possible effect may be that teachers will revise their reading programs to match the content of the test. If the revisions are the result of careful examination of the strengths and weaknesses of a particular curriculum, then the changes will strengthen the entire reading program. However, if the changes result in teaching to the test, then they will have a deleterious effect on the reading program. Hopefully, statewide testing will create an atmosphere which will foster local discussion of the reading process, the skills that are considered integral to that process, and the test's adequacy as a measure of those skills. With a detailed concept of the necessary skill areas and their interrelationships, local administrators and teachers will be able to provide information that will improve the test and thereby make the results more relevant.

Excerpts from

READING AS REASONING

Robert L. Thorndike

> Reading is a very elaborate procedure, involving a weighing of each of many elements in a sentence, their organization in the proper relations one to another, the selection of certain of their connotations and the rejection of others, and the cooperation of many forces to determine final responses. In fact . . . the act of answering simple questions about a simple paragraph . . . includes all the features characteristic of typical reasoning. (Edward L. Thorndike, June 1917, p. 323)

These words appeared 54 years ago in an article by Edward L. Thorndike bearing the same title as this one. The present article offers some lines of evidence that have developed in the intervening half century that lend support and substance to this formulation of the basic nature of the reading process, as that process moves on beyond the initial level at which the crucial task is decoding visually presented symbols to where the crucial task is effective comprehension and response. Although there are many aspects of reading research, this article will concentrate on 3. These are 1) studies of the relationships among different reading skills and of reading skills with measures of abstract intelligence, 2) comparisons of measures of intelligence and of reading as predictors of later academic accomplishments, and 3) cross-national and cross-language studies of reading test performance.

. . . . . .

When we look at the kinds of responses a person is expected to make to a segment of reading material, we are likely to be impressed by their diversity, at least at a superficial level. The reader is expected to identify and comprehend the main idea in a passage, to find the answers to specific questions, to apply the statements in the passage to the solution of problems not directly set forth in the passage, to sense the writer's point of view and purpose in writing, and to recognize the devices by which the mood and flavor of the

From: Reading Research Quarterly, Number 2, 1973-74.

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

International Reading Association

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

passage are maintained. And, of course, in order to do these things, he is expected to know or figure out the meanings of the specific words that the author uses. Reading tests that purport to be "diagnostic" and which are intended to appraise an assortment of these allegedly distinct reading skills have been devised. To what extent are these abilities differentiable? To what extent are different ones of them distinguishable from "reasoning" more generally conceived?

Although a number of studies that analyze the relationships among performance on different reading tasks can be found, all of them suffer from some deficiency and only a couple are technically satisfactory. Frederick B. Davis (1972) has pointed out that the key requirements of such research are that the tasks on the several subtests used be meticulously prepared so that they correspond precisely to the different hypothesized reading skills, and that the analysis account for all the reliably measured variance. It may be added that while the analysis should account for reliable variance that is specific to a single test, as well as that which is general, the analysis should, at the same time, sedulously avoid any attempt to interpret variance that arises from sheer measurement error.

By all odds, the most thoughtfully planned and meticulously gathered set of data is that reported by Davis (1967, 1972) and described in his "Psychometric Research on Comprehension in Reading." Following up on his early (1941) rational and statistical analysis of reading skills, Davis selected, from a very large item pool, test items to measure 8 hypothesized reading skills. The experimental independence of the items was assured by basing each item on a separate bit of reading. Preliminary item analyses were carried out to identify and retain subsets of items that were homogeneous and distinct in the sense that each item correlated more highly with total score for the group of items designed to represent its own skill than it did with total score for any of the 7 other skills. Thus, both editorial and statistical efforts were concentrated on differentiating the several skills. Estimates of the reliability of each of the subtests were obtained with great care. Davis was concerned primarily with analyses to determine the amount of uniqueness in each test--the proportion of variance in each that was non-chance but which was still not predictable from a combination of the remaining tests. At the same time, Davis' data lend themselves extraordinarily well to study by factor analytic procedures--and he has, indeed, factor analyzed them. However, he made an odd shift in his logic as he proceeded from the uniqueness analysis to the factor analysis, factor analyzing the error variance as well as the reliably measured variance. I have taken the liberty of re-analyzing Davis' correlation matrices, using reliability coefficients rather than unities as diagonal entries, and thus limiting the analysis to the dependable variance in each test. The results are shown in Table 1.

Table 1 Factor analysis of Davis Reading Test data (2 parallel matrices)

Part 1: Unrotated factor loadings

Test	Factor 1 A	Factor 1 B	Factor 2 A	Factor 2 B	Factor 3 A	Factor 3 B	h^2 A	h^2 B	r_{11}
1	71	64	08	32	25	20	57	55	58
2	79	77	03	14	-04	04	63	61	64
3	77	77	05	-13	-07	04	60	62	62
4	80	80	13	-09	04	02	65	64	64
5	73	73	25	-14	-18	-18	63	59	59
6	77	78	-07	07	12	-21	61	66	63
7	78	84	-22	02	02	-04	66	71	66
8	77	82	-22	-17	-15	12	67	71	68
% of total variance	58.6	59.4	2.4	2.5	1.7	1.7	62.7	63.6	(63.2)
% of reliable variance	92.7	94.0	3.8	4.0	2.7	2.7	99.4	100.7	

Part 2: Rotated factor loadings of .20 or over

Test	Factor 1 A	Factor 1 B	Factor 2 A	Factor 2 B	Factor 3 A	Factor 3 B
1	76	74	--	--	--	--
2	74	72	22	28	21	--
3	70	60	21	51	24	--
4	79	63	--	49	22	--
5	65	48	--	54	44	25
6	75	62	21	39	--	35
7	71	70	38	44	--	--
8	66	64	46	55	--	--

Designation of tests

1. Remembering word meanings.
2. Inferring word meanings from context.
3. Understanding content stated explicitly.
4. Weaving ideas in the content.
5. Making inferences about the content.
6. Recognizing the author's tone, mood, and purpose.
7. Identifying the author's literary techniques.
8. Following the structure of the content.

Davis divided his test materials into 2 forms, given on different days, and each subtest administered on one day was correlated with the remaining subtests given on the other day. This made his intercorrelations directly comparable with his alternate-form reliability coefficients; it also provided 2 equivalent correlation matrices. The factor analyses for both are shown in Table 1.

Let me call attention to a few salient features of this table. Looking first at the unrotated loadings, one sees that 3 factors completely exhaust the non-chance variance in this set of tests. In the average of the 2 analyses, the 3 factors account for exactly 100 per cent of the reliably measured individual differences. There is nothing left. The first factor by itself accounts for over 93 per cent of the non-chance variance. Thus, the common core underlying this group of developed skills is very large indeed. In his reporting, Davis has focussed his attention on the unique variance, which does in fact exist as seen by the non-chance variance remaining in the second and third factors. But, from the reverse point of view, we can be impressed by the very great communality in the abilities that these reading tasks call forth.

How shall we interpret the 3 factors? I have tried various rotations of the factors represented in this set of tests, using what a colleague would call "orthogonal eye-ball-max" techniques of rotation. The rotations all bring out the fact, demonstrated by Davis in the uniqueness analysis, that the word knowledge test is to some degree distinct from the rest of the reading tests, or conversely, that the remainder of the tests are to some degree distinct from word knowledge. This latter interpretation is what is displayed in the rotated loadings in Table 1. The first factor has been projected through the test of word meanings, and the second and third factors rotated to achieve positive manifold and simple structure in what is left. Beyond the separation of word knowledge, I must confess that I have been unable to make much sense of the patterning of the loadings--in part because the pattern seems to be quite inconsistent as between the 2 samples.

I have played with several other sets of data for which reliabilities are less adequately estimated. . . . The point . . . is that 80 to 90 per cent of the meaningful (that is, non-chance) variance appears in the first factor and that the remaining meaningful variance is almost completely exhausted by 2 (or possibly 3) more factors. However, it has been difficult to arrive at a consistent interpretation of factors beyond the first, and where there are replications of the same set of variables with different samples, as in Margaret M. Conant's (1942) data, the factor patterns of the later factors are erratic and replication is poor.

. . . . . .

Davis would, I believe, call Factor 1 of Table 1 "word knowledge," and Factor 2 and/or 3 "reasoning in reading." But wait a minute! Is the ability that is measured by a word knowledge test reasoning-free? Remember, for example, that the vocabulary subtest of the Stanford-Binet has higher correlation with Binet IQ than any other subtest.

A somewhat different correlational analysis of word knowledge and paragraph reading materials was afforded a while back, when we had the opportunity to transcribe and analyze the standardized test results for a suburban community lying not far from New York City. This community had given an intelligence test in grades 1, 3, 5, and 7 and an achievement battery in grades 2, 4, 6, and 8. For the present, we are concerned only with the reading tests from the achievement battery. I have factor analyzed the correlations of the 4 intelligence tests and the 4 reading tests. . . .

Once again, the first factor is a swamping one, accounting for more than 90 per cent of the common-factor variance. If reading isn't reasoning, then maybe reasoning is reading! The rotation of factors has been carried out so that the first factor is defined by the early intelligence tests. But a first factor, so defined, correlates substantially and about equally with the word knowledge and paragraph reading scores in later tests. To draw a sharp contrast between "word knowledge" and "reasoning in reading" seems hardly justified by these results. If intelligence tests assess reasoning (and this can of course be debated), word knowledge tests seem to be as much an indicator of that reasoning as do questions based on connected prose; and both seem to serve this function admirably.

. . . . . .

There is another angle from which one may view this same set of aptitude and reading achievement data. One can view the successive measures as predictors of terminal status--either terminal scholastic aptitude or terminal academic achievement. . . .

. . . . . .

. . . With one or 2 minor exceptions, the reading tests and the aptitude tests make about equal contributions, whether the end result to be predicted is aptitude, reading, or general achievement.

We can also look at the weights of the several predictors, when all are working together in a regression equation. . . . Here we see a little difference between the aptitude and the achievement measures. Possibly reading is slightly more powerful as a predictor of reading than of other outcomes. Prediction of final achievement can be accomplished very largely by the last previous cycle of aptitude and reading measures, and these appear to incorporate most of what

has gone before. It is as if each new layer of competence was superimposed upon and incorporated all previous layers. But for the seventh grade aptitude measures, all but one of the earlier measures does have some unique contribution to make.

Our results . . . suggest that measures of reading should be considered as predictors of academic achievement fully on a par with tests to which the term "scholastic aptitude" is conventionally applied. This conclusion is affirmed even more strongly in some (unpublished) data supplied by Professor Joseph Breen of St. Johns University. Breen followed up groups originally tested in first or second grade and got final measures with the Iowa Tests of Educational Development in grade 10 or 11. . . .

. . . . . .

. . . The best predictors of final achievement are in each instance scores on an earlier reading test. . . . Almost without exception the regression weights for reading are larger than those for IQ.

Why is it that a measure of ability to construe the printed page provides such a powerful indicator of skills intellectual and academic? One cue comes from cross-national and cross-language studies of educational achievement in which, with a large group of associates, I have been participating for the past decade. One of the early studies in this series (A. W. Foshay et al., 1962) included, among other tests, a 30-item test of reading comprehension. With a good deal of trepidation, because the problems of translation into 8 languages had seemed particularly horrendous for a reading test, we embarked upon the analysis of the results from this test.

One analysis consisted of an examination of the between-countries variance, in comparison with the average within-country variance. The per cent attributed to national differences for the reading test was about the smallest of any of the tests we studied, and noticeably less than for arithmetic or for a presumably "culture-free" non-verbal aptitude test. In our data, the reading test appeared to be the closest approximation to an internationally culture-fair measure. A second analysis was a cross-national correlation of single-item difficulties, to determine whether the same items were hard and easy in different countries and in different languages. . . .

. . . there are no correlations below 0.80. The average correlation, excluding pairs of countries in which the same language is spoken, is 0.86. An item is an item is an item, whether in English, French, German, Swedish, Polish, Finnish, Hebrew, or Serbo-Croatian!

The difficulty lies, it appears, not in the verbiage or syntax of a specific language, but in the structure of ideas with which the item deals. At age 13, when these youngsters were tested, reading is no longer--to any substantial degree--a decoding problem, these results would seem to say. It is a thinking problem.

. . . . . .

[These] results . . . were obtained at a single age level (13) in a group of developed occidental countries. How would results compare at other ages or with other national groups? In a Teachers College dissertation, Kumbaraci (1966) found a correlation of about 0.70 for Turkish and American students in the final years of secondary school. We have recently reported results of a survey of reading comprehension at ages 10, 14, and end of secondary school in some 15 countries, including regions in India, Iran, and Chile (R. L. Thorndike, 1973). The results suggest that inclusion of oriental and of less developed countries has resulted in rather lower correlations for item difficulties across countries, so that they correspond more nearly to the 0.70 found by Kumbaraci. Is this because of wider linguistic differences, or because of a greater diversity of cultural backgrounds that makes a given idea more familiar, and consequently simpler in one culture than another? Is it that some occidental concepts have no cognates in oriental languages? As of now, I haven't even the evidence to permit a shrewd guess. But in these cross-national studies we see further evidence that the "thinking" component as distinct from the "decoding" component seems to bulk large in the task called reading.

It has been suggested here that performance on a wide variety of reading tasks calls primarily for some one common ability; that this ability can be distinguished only with difficulty from that represented in general measures of verbal intelligence or scholastic aptitude, and may be thought of as verbal reasoning; that level of early competence in reading is a very potent predictor of eventual level of academic performance in the conventional word-centered curriculum. As further evidence of the reasoning component in reading tasks, I have pointed to the apparent cross-national and cross-linguistic stability of reading tasks. What might all this mean for education?

If reading is reasoning, we face at one and the same time a barrier and a challenge. The barrier is that set by the child's limited comprehension of what he reads, which we see now as not primarily a deficit in one or more specific and readily teachable reading skills but as a reflection of generally meager intellectual processes. And this barrier promises to stand in the way of a wide range of future learnings.

The challenge is to overcome this barrier by better and

more inventive teaching--not solely to read, but also to think. Because as we improve the understanding with which a child reads, we may concurrently improve the effectiveness with which he processes a wide range of information important in his development. The challenge is also to learn to exploit for educational advantage the individual's resources for reasoning through other media than words, so that the barrier of verbal limitation may be bypassed whenever it is not relevant.

REFERENCES

Conant, Margaret M. The construction of a diagnostic reading test. New York: Bureau of Publications, Teachers College, Columbia University, 1942.

Davis, Frederick B. Identification and measurement of reading skills of high school students. USOE Cooperative Research Project No. 3023. Philadelphia: University of Pennsylvania, 1967.

Davis, Frederick B. Psychometric research on comprehension in reading. Reading Research Quarterly, Summer 1972. 7, 628-678.

Foshay, A. W., et al. Educational achievements of thirteen-year-olds in twelve countries. Hamburg, Germany: UNESCO Institute for Education, 1962.

Kumbaraci, Turkan E. Translated reading tests as culture-fair measures for foreign students. Unpublished doctoral dissertation. New York: Teachers College, Columbia University, 1966.

Thorndike, Edward L. Reading as reasoning: a study of mistakes in paragraph reading. Journal of Educational Psychology, June 1917, 8, 323-332.

Thorndike, Robert L. Reading comprehension education in fifteen countries. International Studies in Evaluation III. New York: Halsted Press, 1973.

TEST REFERENCE

The Iowa Tests of Educational Development. E. F. Lindquist & Leonard S. Feldt. Chicago: Science Research Associates, 1942-63.

Editor's Note: Interested readers are encouraged to read the original article which contains additional data to illustrate the author's points.

11:00 a.m. COFFEE

11:20 a.m. SECTIONS

SECTION I - ELEMENTARY SCHOOLS (4th grade test)

Chairman: Dr. Edward Fry
Panel Member: Dr. Raymond Arciszewski, Principal
Southern Blvd. School, Chatham Township
Resource Person: Dr. Roy Kress

SECTION II - MIDDLE SCHOOLS (7th grade test)

Chairman: Dr. Evelyn Slobodzian
Panel Member: Dr. George Phillips, Queens College (CUNY)
Resource Person: Ms. Joan Fischer, Rutgers University
Graduate School of Education

SECTION III - HIGH SCHOOLS (12th grade test)

Chairman: Dr. Joseph Zelnick, Livingston College
Panel Member: Dr. Josephine Goldsmith, Rutgers College
Resource Person: Dr. Henry S. Dyer

SECTION IV - ADMINISTRATION

Chairman: Dr. Lawrence Kaplan, Rutgers
Graduate School of Education
Panel Member: Dr. Joseph A. Kreskey
Asst. Supt., Edison Public Schools
Resource Person: Dr. Trika Smith-Burke, New York University

SECTION V - SPECIALISTS (Reading and Testing Personnel, School Psychologists, Learning Disability Specialists, Counselors, Curriculum Specialists)

Chairman: Dr. Martin Kling, Rutgers University
Graduate School of Education
Panel Member: Mrs. Roslyn Gross, Director
Pupil Personnel Services and Evaluation
Woodbridge Public Schools
Resource Person: Dr. Robert Thorndike

1:00 p.m. LUNCH (Commons)

2:00 p.m. GENERAL SESSION - Section Chairmen give brief report and ask a majority question - General Session Speakers respond.

3:00 p.m. ADJOURN